

# Maximum entropy distributions on graphs

Christopher J. Hillar\*

Andre Wibisono†

University of California, Berkeley

January 16, 2013

## Abstract

Inspired by the problem of sensory coding in neuroscience, we study the maximum entropy distribution on weighted graphs with a given expected degree sequence. This distribution on graphs is characterized by independent edge weights parameterized by vertex potentials at each node. Using the general theory of exponential family distributions, we prove existence and uniqueness of the maximum likelihood estimator (MLE) of the vertex parameters. We also prove in several cases the surprising consistency of the MLE from a single graph sample, extending results of Chatterjee, Diaconis, and Sly for unweighted (binary) graphs. Interestingly, our extensions require an intricate study of the inverses of diagonally dominant positive matrices. Along the way, we derive analogues of the Erdős-Rényi criterion of graphic sequences for weighted graphs.

**Notation:** We use the notation  $\mathbb{R}_+ = (0, \infty)$ ,  $\mathbb{R}_0 = [0, \infty)$ ,  $\mathbb{N} = \{1, 2, \dots\}$ , and  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ . We write  $\sum_{(i,j)}$  and  $\prod_{(i,j)}$  for the summation and product, respectively, over all  $\binom{n}{2}$  pairs  $(i, j)$  with  $i \neq j$ . Given a subset  $C$  of  $\mathbb{R}^n$ , we let  $C^\circ$  and  $\overline{C}$  denote the interior and closure of  $C$  in  $\mathbb{R}^n$ , respectively. For a vector  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ , we set  $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$  to be the  $\ell_\infty$ -norm of  $x$ .

## 1 Introduction

Maximum entropy models are an important class of statistical models for biology. For instance, they have been found recently to well-model protein folding [30, 36], antibody diversity [25], neural population activity [32, 34, 38, 37, 4, 42, 35], and flock behavior [5]. Here, we develop a general framework for studying maximum entropy distributions on weighted graphs, extending recent work of Chatterjee, Diaconis, and Sly [8]. Our motivation for developing this theory comes from the problem of sensory coding in neuroscience.

In the brain, information is represented by discrete electrical pulses, called *action potentials* or *spikes* [29]. This includes neural representations of sensory stimuli which can take on a continuum of values. For instance, large photoreceptor arrays in the retina respond to a range of light intensities in a visual environment, but the brain does not receive information from these photoreceptors directly. Instead, retinal ganglion cells must convey this detailed input to the visual cortex using only a series of binary electrical signals. Continuous stimuli are therefore converted by networks of neurons to sequences of spike times.

An unresolved controversy in neuroscience is whether information is contained in the precise timings of these spikes or only in their “rates” (i.e., counts of spikes in a window of time). Early theoretical studies [23] suggest that information capacities of timing-based codes are superior to those that are rate-based (also see [17] for an implementation in a simple model). Moreover, a number of scientific articles have appeared

---

\*Redwood Center for Theoretical Neuroscience, [chillar@msri.org](mailto:chillar@msri.org); partially supported by National Science Foundation Grant IIS-0917342 and an National Science Foundation All-Institutes Postdoctoral Fellowship administered by the Mathematical Sciences Research Institute through its core grant DMS-0441170.

†Department of Electrical Engineering and Computer Science, [wibisono@eecs.berkeley.edu](mailto:wibisono@eecs.berkeley.edu).

suggesting that precise spike timing [1, 3, 27, 40, 22, 6, 24, 26, 11, 19] and synchrony [39] are important for various computations in the brain.<sup>1</sup> Here, we briefly explain a possible scheme for encoding continuous vectors with spiking neurons that takes advantage of precise spike timing and the mathematics of maximum entropy distributions. A more detailed examination of this model will appear in a future work.

Consider a network of  $n$  neurons in one region of the brain which transmit a continuous vector  $\theta \in \mathbb{R}^n$  using sequences of spikes to a second receiver region. We assume that this second region contains a number of coincidence detectors that measure the absolute difference in spike times between pairs of neurons projecting from the first region. We imagine three scenarios for how information can be obtained by these detectors. In the first, the detector is only measuring for synchrony between spikes; that is, either the detector assigns a 0 to a nonzero timing difference or a 1 to a coincidence of spikes. In another scenario, timing differences between projecting neurons can assume an infinite but countable number of possible values. Finally, in the third situation, we allow these differences to take on any nonnegative real value. We further assume that neuronal output and thus spike times are stochastic variables. A basic question now arises: How can the first region encode  $\theta$  so that it can be recovered robustly by the second?

We answer this question by first asking the one symmetric to this: How can the second region recover a real vector transmitted by an unknown sender region from spike timing measurements? We propose the following solution to this problem. Fix one of the detector mechanics as described above, and set  $a_{ij}$  to be the measurement of the absolute timing difference between spikes from projecting neurons  $i$  and  $j$ . We assume that the receiver population can compute the (local) sums  $\hat{d}_i = \sum_{j \neq i} a_{ij}$  efficiently. The values  $\mathbf{a} = (a_{ij})$  represent a weighted graph  $G$  on  $n$  vertices, and we assume that  $a_{ij}$  is randomly drawn from a distribution on timing measurements  $(A_{ij})$ . Making no further assumptions, a principle of Jaynes [18] suggests that the second region propose that the timing differences are drawn from the (unique) distribution over weighted graphs with the most entropy [33, 10] having the vector  $\hat{\mathbf{d}} = (\hat{d}_1, \dots, \hat{d}_n)$  for the expectations  $\mathbb{E}[\sum_{j \neq i} A_{ij}]$  of the degree sums  $\sum_{j \neq i} A_{ij}$ . Depending on which of the three scenarios described above is true for the coincidence detector, this prescription produces one of three different maximum entropy distributions.

Consider the third scenario above (the other cases are also subsumed by our results). As we shall see in Section 3.2, the distribution determined in this case is parameterized by a real vector  $\theta = (\theta_1, \dots, \theta_n)$ , and the maximum likelihood estimator (MLE) for these parameters using  $\hat{\mathbf{d}}$  as sufficient statistics boils down to solving the following set of  $n$  algebraic equations in the  $n$  unknowns  $\theta_1, \dots, \theta_n$ :

$$\hat{d}_i = \sum_{j \neq i} \frac{1}{\theta_i + \theta_j} \quad \text{for } i = 1, \dots, n. \quad (1)$$

Given our motivation, we call the system of equations (1) the *retina equations* for theoretical neuroscience, and note that they have been studied in a more general context by Sanyal, Sturmfels, and Vinzant [31] using matroid theory and algebraic geometry. Somewhat remarkably, a solution  $\hat{\theta}$  to (1) has the property that it is arbitrarily close to the original parameters  $\theta$  for sufficiently large network sizes  $n$  (in the scenario of binary measurements, this is a result of [8]). In particular, it is possible for the receiver region to recover reliably a continuous vector  $\theta$  from a *single* cycle of neuronal firing emanating from the sender region.

We now know how to answer our first question: *The sender region should arrange spike timing differences to come from a maximum entropy distribution.* We remark that this conclusion is consistent with modern paradigms in artificial intelligence, such as the concept of the Boltzmann machine [2], which is a stochastic version of its (zero-temperature) deterministic limit, the Little-Hopfield network [21, 16].

The organization of this paper is as follows. In Section 2, we lay out the basic theory of maximum entropy distributions on graphs. Section 3 is devoted to specializing the theory to three common weight sets and contains an extension of the Erdős-Rényi graphic sequence criterion for weighted graphs. In Section 4, we prove the consistency of the MLE from a single sample in the examples from Section 3. A key step in our proofs is a new inequality for the norm of the inverses of positive, symmetric diagonally dominant matrices from [15]. Finally, Appendix A contains some facts about subexponential random variables.

---

<sup>1</sup>Although it is well-known that precise spike timing is used for time-disparity computation in animals [7], such as when owls track prey with binocular hearing or when electric fish use electric fields around their bodies for locating objects.

## 2 General theory via exponential family distributions

In this section we develop the general machinery of maximum entropy distributions on graphs via the theory of exponential family distributions [41], and in subsequent sections we analyze particular cases. Consider an undirected graph  $G$  on  $n \geq 2$  vertices with edge  $(i, j)$  having weight  $a_{ij} \in S$ , where  $S \subseteq \mathbb{R}$  is the set of possible weight values. We will later consider the specific cases  $S = \{0, 1\}$  (unweighted graphs),  $S = \mathbb{R}_0$  (weighted graphs with continuous weights), and  $S = \mathbb{N}_0$  (weighted graphs with discrete weights). A graph  $G$  is fully specified by its *adjacency matrix*  $\mathbf{a} = (a_{ij})_{i,j=1}^n$ , which is an  $n \times n$  symmetric matrix with zeros along its diagonal. A probability distribution over graphs  $G$  corresponds to a distribution over adjacency matrices  $\mathbf{a} = (a_{ij}) \in S^{(n)}$ . Given a graph with adjacency matrix  $\mathbf{a} = (a_{ij})$ , let  $\deg_i(\mathbf{a}) = \sum_{j \neq i} a_{ij}$  be the degree of vertex  $i$ , and let  $\deg(\mathbf{a}) = (\deg_1(\mathbf{a}), \dots, \deg_n(\mathbf{a}))$  be the degree sequence of  $\mathbf{a}$ .

Let  $\mathcal{S}$  be a  $\sigma$ -algebra over the set of weight values  $S$ . Assume there is a canonical  $\sigma$ -finite probability measure  $\nu$  on  $(S, \mathcal{S})$ . Let  $\nu^{(n)}$  be the product measure on  $S^{(n)}$ . Let  $\mathfrak{P}$  be the set of all probability distributions on  $S^{(n)}$  that are absolutely continuous with respect to  $\nu^{(n)}$ . Since  $\nu^{(n)}$  is  $\sigma$ -finite, these probability distributions can be characterized by their density functions, i.e. the Radon-Nikodym derivatives with respect to  $\nu^{(n)}$ . Given a sequence  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}^n$ , let  $\mathfrak{P}_{\mathbf{d}}$  be the set of distributions in  $\mathfrak{P}$  whose expected degree sequence is equal to  $\mathbf{d}$ ,

$$\mathfrak{P}_{\mathbf{d}} = \{\mathbb{P} \in \mathfrak{P} : \mathbb{E}_{\mathbb{P}}[\deg(A)] = \mathbf{d}\},$$

where in the definition above, the random variable  $A = (A_{ij}) \in S^{(n)}$  is drawn from the distribution  $\mathbb{P}$ . Then the distribution  $\mathbb{P}^*$  in  $\mathfrak{P}_{\mathbf{d}}$  with maximum entropy is the exponential family distribution with the degree sequence as sufficient statistics [41, Chapter 3]. That is, the density of  $\mathbb{P}^*$  at  $\mathbf{a} = (a_{ij}) \in S^{(n)}$  is given by<sup>2</sup>

$$p^*(\mathbf{a}) = \exp(-\theta^\top \deg(\mathbf{a}) - Z(\theta)), \quad (2)$$

where  $Z(\theta)$  is the *log-partition function*,

$$Z(\theta) = \log \int_{S^{(n)}} \exp(-\theta^\top \deg(\mathbf{a})) \nu^{(n)}(d\mathbf{a}),$$

and  $\theta = (\theta_1, \dots, \theta_n)$  belongs to the *natural parameter space*

$$\Theta = \{\theta \in \mathbb{R}^n : Z(\theta) < \infty\}.$$

Recalling that  $\deg_i(\mathbf{a}) = \sum_{j \neq i} a_{ij}$ , we can write

$$\exp(-\theta^\top \deg(\mathbf{a})) = \exp\left(-\sum_{i=1}^n \sum_{j \neq i} \theta_i a_{ij}\right) = \exp\left(-\sum_{(i,j)} (\theta_i + \theta_j) a_{ij}\right) = \prod_{(i,j)} \exp(-(\theta_i + \theta_j) a_{ij}).$$

Hence, we can express the log-partition function as

$$Z(\theta) = \log \prod_{(i,j)} \int_S \exp(-(\theta_i + \theta_j) a_{ij}) \nu(da_{ij}) = \sum_{(i,j)} Z_1(\theta_i + \theta_j),$$

in which  $Z_1(t)$  is the marginal log-partition function

$$Z_1(t) = \log \int_S \exp(-ta) \nu(da).$$

Consequently, the density in (2) can be written as

$$p^*(\mathbf{a}) = \prod_{(i,j)} \exp(-(\theta_i + \theta_j) a_{ij} - Z_1(\theta_i + \theta_j)),$$

---

<sup>2</sup>We choose to use  $-\theta$  in the parameterization (2), and not the canonical parameterization  $p^*(\mathbf{a}) \propto \exp(\theta^\top \deg(\mathbf{a}))$ , because it simplifies the notations in our later presentation.

from which we see that the edge weights  $A_{ij}$  are independent random variables, with  $A_{ij} \in S$  having distribution  $\mathbb{P}_{ij}^*$  with density

$$p_{ij}^*(a) = \exp \left( -(\theta_i + \theta_j)a - Z_1(\theta_i + \theta_j) \right).$$

In particular, the edge weights  $A_{ij}$  belong to the same exponential family distribution but with different parameters that depend on  $\theta_i$  and  $\theta_j$  (or rather, on their sum  $\theta_i + \theta_j$ ). The parameters  $\theta_1, \dots, \theta_n$  can be interpreted as the potential at each vertex that determines how strongly the vertices are connected to each other. Furthermore, we can write the natural parameter space  $\Theta$  as

$$\Theta = \{\theta \in \mathbb{R}^n : Z_1(\theta_i + \theta_j) < \infty \text{ for all } i \neq j\}.$$

Going back to the characterization of  $\mathbb{P}^*$  as the maximum entropy distribution in  $\mathfrak{P}_{\mathbf{d}}$ , the condition  $\mathbb{P}^* \in \mathfrak{P}_{\mathbf{d}}$  means that we choose the parameter  $\theta$  such that  $\mathbb{E}_{\mathbb{P}^*}[\deg(A)] = \mathbf{d}$ . Equivalently, noting that  $-\nabla Z(\theta) = \mathbb{E}_{\mathbb{P}^*}[\deg(A)]$ , the solution to  $-\nabla Z(\theta) = \mathbf{d}$  is precisely the maximum likelihood estimator (MLE) of  $\theta$  given an empirical degree sequence  $\mathbf{d} \in \mathbb{R}^n$ . For instance, the vector  $\mathbf{d}$  can be the average of the degree sequences of graphs  $G_1, \dots, G_m$  drawn i.i.d. from the distribution  $\mathbb{P}^*$ . We will later study the properties of the MLE of  $\theta$  from a *single* sample  $G \sim \mathbb{P}^*$ . For now, we address the question regarding the existence and uniqueness of this MLE assuming the degree sequence  $\mathbf{d}$  is given.

Define the *mean parameter space*  $\mathcal{M}$  to be the set of expected degree sequences from all distributions on  $S^{\binom{n}{2}}$  that are absolutely continuous with respect to  $\nu^{\binom{n}{2}}$ :

$$\mathcal{M} = \{\mathbb{E}_{\mathbb{P}}[\deg(A)] : \mathbb{P} \in \mathfrak{P}\}.$$

The set  $\mathcal{M}$  is necessarily convex, since a convex combination of probability distributions in  $\mathfrak{P}$  is also a probability distribution in  $\mathfrak{P}$ . Recall that an exponential family distribution is *minimal* if there is no linear combination of the sufficient statistics that is constant almost surely with respect to the base distribution. This is true for  $\mathbb{P}^*$ , for which the sufficient statistics are the degree sequence. We say that  $\mathbb{P}^*$  is *regular* if the natural parameter space  $\Theta$  is open. By the general theory of exponential family distributions [41, Theorem 3.3], in a regular and minimal exponential family distribution, the gradient of the log-partition function maps the natural parameter space  $\Theta$  to the interior of the mean parameter space  $\mathcal{M}$ , and this mapping<sup>3</sup>

$$-\nabla Z : \Theta \rightarrow \mathcal{M}^\circ$$

is bijective. Thus we have established the following.

**Proposition 2.1.** *Assume  $\Theta$  is open. Then there exists a solution  $\theta \in \Theta$  to the MLE equation  $\mathbb{E}_{\mathbb{P}^*}[\deg(A)] = \mathbf{d}$  if and only if  $\mathbf{d} \in \mathcal{M}^\circ$ , and if such a solution exists then it is unique.*

It remains to characterize the mean parameter space  $\mathcal{M}$ . We say that a sequence  $\mathbf{d} = (d_1, \dots, d_n)$  is *graphic* if  $\mathbf{d}$  is the degree sequence of a graph  $G$  with edge weights in  $S$ , and in this case we say that  $G$  *realizes*  $\mathbf{d}$ . It is important to note that whether a sequence  $\mathbf{d}$  is graphic depends on the weight set  $S$ , which we fix for now. Let  $\mathcal{W}$  be the set of all graphic sequences, and let  $\text{conv}(\mathcal{W})$  be the convex hull of  $\mathcal{W}$ . Clearly we have  $\mathcal{M} \subseteq \text{conv}(\mathcal{W})$ , because any element of  $\mathcal{M}$  is of the form  $\mathbb{E}_{\mathbb{P}}[\deg(A)]$  for some distribution  $\mathbb{P}$  and  $\deg(A) \in \mathcal{W}$  for every realization of the random variable  $A$ . On the other hand, suppose  $\mathfrak{P}$  contains the Dirac delta measures  $\delta_B$  for each  $B \in S^{\binom{n}{2}}$ . Given  $\mathbf{d} \in \mathcal{W}$ , let  $B$  be the adjacency matrix of the graph that realizes  $\mathbf{d}$ . Then  $\mathbf{d} = \mathbb{E}_{\delta_B}[\deg(A)] \in \mathcal{M}$ , which means  $\mathcal{W} \subseteq \mathcal{M}$ , and hence  $\text{conv}(\mathcal{W}) \subseteq \mathcal{M}$  since  $\mathcal{M}$  is convex. Thus in this case we have  $\mathcal{M} = \text{conv}(\mathcal{W})$ . However, in general  $\mathfrak{P}$  might not contain Dirac measures, and we need to look at the specific structure of  $\mathfrak{P}$  to decide whether  $\text{conv}(\mathcal{W}) \subseteq \mathcal{M}$ .

We emphasize the distinction between a *valid* solution  $\theta \in \Theta$  and a *general* solution  $\theta \in \mathbb{R}^n$  to the MLE equation  $\mathbb{E}_{\mathbb{P}^*}[\deg(A)] = \mathbf{d}$ . As we saw from Proposition 2.1, we have a precise characterization of the existence and uniqueness of the valid solution  $\theta \in \Theta$ , but in general, there are multiple solutions  $\theta$  to the

<sup>3</sup>The presence of the minus sign in the mapping  $-\nabla Z$  is due to our choice of the parameterization in (2) using  $-\theta$ .

MLE equation. In this paper we shall be concerned only with the valid solution; Sanyal, Sturmfels, and Vinzant study some algebraic properties of general solutions [31].

We close this section by discussing the symmetry of the valid solution to the MLE equation. Let  $\text{Dom}(Z_1) = \{t \in \mathbb{R} : Z_1(t) < \infty\}$ , and let  $\mu : \text{Dom}(Z_1) \rightarrow \mathbb{R}$  be the *mean function*

$$\mu(t) = \int_S a \exp(-ta - Z_1(t)) \nu(da).$$

Observing that we can write

$$\mathbb{E}_{\mathbb{P}^*}[A_{ij}] = \int_S a \exp(-(\theta_i + \theta_j)a - Z_1(\theta_i + \theta_j)) \nu(da) = \mu(\theta_i + \theta_j),$$

the MLE equation  $\mathbb{E}_{\mathbb{P}^*}[\deg(A)] = \mathbf{d}$  then becomes

$$d_i = \sum_{j \neq i} \mu(\theta_i + \theta_j) \quad \text{for } i = 1, \dots, n. \quad (3)$$

In the statement below,  $\text{sgn}$  denotes the sign function:  $\text{sgn}(t) = t/|t|$  if  $t \neq 0$ , and  $\text{sgn}(0) = 0$ .

**Proposition 2.2.** *Let  $\mathbf{d} \in \mathcal{M}^\circ$ , and let  $\theta \in \Theta$  be the unique solution to the system of equations (3). If  $\mu$  is strictly increasing, then*

$$\text{sgn}(d_i - d_j) = \text{sgn}(\theta_i - \theta_j) \quad \text{for all } i \neq j,$$

*and similarly, if  $\mu$  is strictly decreasing, then*

$$\text{sgn}(d_i - d_j) = \text{sgn}(\theta_j - \theta_i) \quad \text{for all } i \neq j.$$

*Proof.* Given  $i \neq j$ ,

$$\begin{aligned} d_i - d_j &= \left( \mu(\theta_i + \theta_j) + \sum_{k \neq i, j} \mu(\theta_i + \theta_k) \right) - \left( \mu(\theta_j + \theta_i) + \sum_{k \neq i, j} \mu(\theta_j + \theta_k) \right) \\ &= \sum_{k \neq i, j} (\mu(\theta_i + \theta_k) - \mu(\theta_j + \theta_k)). \end{aligned}$$

If  $\mu$  is strictly increasing, then  $\mu(\theta_i + \theta_k) - \mu(\theta_j + \theta_k)$  has the same sign as  $\theta_i - \theta_j$  for each  $k \neq i, j$ , and thus  $d_i - d_j$  also has the same sign as  $\theta_i - \theta_j$ . Similarly, if  $\mu$  is strictly decreasing, then  $\mu(\theta_i + \theta_k) - \mu(\theta_j + \theta_k)$  has the opposite sign of  $\theta_i - \theta_j$ , and thus  $d_i - d_j$  also has the opposite sign of  $\theta_i - \theta_j$ .  $\square$

### 3 Specific edge weight cases

We now consider specific choices of the weight set  $S$ . In each case we investigate the distribution of the edge weights  $A_{ij}$ , the natural parameter space  $\Theta$ , and characterize the mean parameter space  $\mathcal{M}$ .

#### 3.1 Unweighted graphs

Let  $S = \{0, 1\}$  and  $\nu$  be the counting measure so that we consider simple unweighted graphs. In this case,

$$Z_1(t) = \log(1 + \exp(-t)) < \infty \quad \text{for all } t \in \mathbb{R},$$

so  $\text{Dom}(Z_1) = \mathbb{R}$  and the natural parameter space is  $\Theta = \mathbb{R}^n$ , which is open. The edge weights  $A_{ij}$  are then independent Bernoulli random variables with

$$\mathbb{P}^*(A_{ij} = 1) = \frac{\exp(-\theta_i - \theta_j)}{1 + \exp(-\theta_i - \theta_j)} = \frac{1}{1 + \exp(\theta_i + \theta_j)}.$$

This model has been studied recently by Chatterjee, Diaconis, and Sly [8] in the context of graph limits. When  $\theta_1 = \theta_2 = \dots = \theta_n = t$ , we recover the classical Erdős-Rényi model with edge emission probability  $p = 1/(1 + \exp(2t))$ .

Since  $\nu^{\binom{n}{2}}$  is the counting measure on  $\{0, 1\}^{\binom{n}{2}}$ , all distributions on  $\{0, 1\}^{\binom{n}{2}}$  are absolutely continuous with respect to  $\nu^{\binom{n}{2}}$ , so  $\mathfrak{P}$  contains all probability distributions on  $\{0, 1\}^{\binom{n}{2}}$ . In particular,  $\mathfrak{P}$  contains the Dirac measures, and hence by the discussion in the preceding section we have  $\mathcal{M} = \text{conv}(\mathcal{W})$ , where  $\mathcal{W}$  is the set of all graphic sequences. Thus, it now remains to characterize when a given sequence  $\mathbf{d} \in \mathbb{N}_0^n$  is a degree sequence of an unweighted graph, which is precisely what a classical result of Erdős and Gallai tells us.

**Lemma 3.1** (Erdős-Gallai [13]). *A sequence  $(d_1, \dots, d_n) \in \mathbb{N}_0^n$  is graphic if and only if  $\sum_{i=1}^n d_i$  is even and*

$$\sum_{i=1}^k d_i \leq k(k-1) + \sum_{i=k+1}^n \min\{d_i, k\} \quad \text{for } k = 1, \dots, n. \quad (4)$$

The mean function is given by  $\mu(t) = 1/(1 + \exp(t))$ , which is strictly decreasing for all  $t \in \text{Dom}(Z_1) = \mathbb{R}$ . Given  $\mathbf{d} \in \mathbb{R}^n$ , the MLE equation  $\mathbb{E}_{\mathbb{P}^*}[\deg(A)] = \mathbf{d}$  becomes the system of equations

$$d_i = \sum_{j \neq i} \frac{1}{1 + \exp(\theta_i + \theta_j)} \quad \text{for } i = 1, \dots, n, \quad (5)$$

and we want to find a solution  $\theta \in \Theta = \mathbb{R}^n$ . We know that there exists a unique solution  $\theta \in \mathbb{R}^n$  if and only if  $\mathbf{d} \in \text{conv}(\mathcal{W})^\circ$ . However, given the characterization of  $\mathcal{W}$  by the Erdős-Gallai criterion, it is unclear how to decide whether a given  $\mathbf{d}$  belongs to  $\text{conv}(\mathcal{W})^\circ$ . Nevertheless, in practice we can circumvent this issue by employing the following iterative algorithm proposed in [8] to find the MLE solution; thus, given a sequence  $\mathbf{d}$  in practice, we can run this algorithm and see whether it converges. Note that clearly if  $\mathbf{d} \in \mathbb{R}^n$  has  $d_i \leq 0$  for some  $i = 1, \dots, n$  then  $\mathbf{d} \notin \text{conv}(\mathcal{W})^\circ$ , so we focus on the case  $\mathbf{d} \in \mathbb{R}_+^n$ .

**Proposition 3.2** ([8, Theorem 1.5]). *Given  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}_+^n$ , define the function  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^n$  by  $\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \dots, \varphi_n(\mathbf{x}))$ , where*

$$\varphi_i(\mathbf{x}) = -\log d_i + \log \sum_{j \neq i} \frac{1}{\exp(x_j) + \exp(-x_i)} \quad \text{for } i = 1, \dots, n.$$

*Given any  $\mathbf{x}_0 \in \mathbb{R}^n$ , let  $\mathbf{x}_{k+1} = \varphi(\mathbf{x}_k)$  for  $k \in \mathbb{N}_0$ . Suppose  $\mathbf{d} \in \text{conv}(\mathcal{W})^\circ$  and let  $\theta \in \mathbb{R}^n$  be the unique solution to (5). Then there exists a constant  $C$  that depends on  $\|\mathbf{x}_0\|_\infty$  and  $\|\theta\|_\infty$ , such that*

$$\|\mathbf{x}_{k+2} - \theta\|_\infty \leq C \|\mathbf{x}_k - \theta\|_\infty \quad \text{for all } k \in \mathbb{N}_0.$$

*In particular, this means  $(\mathbf{x}_k)$  converges exponentially fast to the MLE solution  $\theta$ . On the other hand, if  $\mathbf{d} \notin \text{conv}(\mathcal{W})^\circ$  then  $(\mathbf{x}_k)$  has a divergent subsequence.*

We summarize the discussion in this section in the following corollary.

**Corollary 3.3.** *Given  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}^n$ , the system of equations*

$$d_i = \sum_{j \neq i} \frac{1}{1 + \exp(\theta_i + \theta_j)} \quad \text{for } i = 1, \dots, n$$

*has a solution  $\theta \in \mathbb{R}^n$  if and only if  $\mathbf{d} \in \text{conv}(\mathcal{W})^\circ$ . Furthermore, if such a solution exists then it is unique and is the fixed point of the iterative algorithm in Proposition 3.2, and it has the property that*

$$\text{sgn}(d_i - d_j) = \text{sgn}(\theta_j - \theta_i) \quad \text{for all } i \neq j.$$

### 3.2 Weighted graphs with continuous weights

Now let  $S = \mathbb{R}_0$  and  $\nu$  be the Lebesgue measure on  $\mathbb{R}_0$ , so we are considering weighted graphs with continuous weights. In this case the marginal log-partition function is

$$Z_1(t) = \log \int_{\mathbb{R}_0} \exp(-ta) da = \begin{cases} \log(1/t) & \text{if } t > 0 \\ \infty & \text{if } t \leq 0. \end{cases}$$

Thus  $\text{Dom}(Z_1) = \mathbb{R}_+$ , and the natural parameter space is

$$\Theta = \{(\theta_1, \dots, \theta_n) \in \mathbb{R}^n : \theta_i + \theta_j > 0 \text{ for } i \neq j\},$$

which is open. For  $\theta \in \Theta$ , the edge weights  $A_{ij}$  are independent exponential random variables with density

$$p_1^*(a) = (\theta_i + \theta_j) \exp(-(\theta_i + \theta_j)a) \quad \text{for } a \in \mathbb{R}_0$$

and mean parameter  $\mathbb{E}_{\mathbb{P}^*}[A_{ij}] = 1/(\theta_i + \theta_j)$ . Thus, the MLE equation  $\mathbb{E}_{\mathbb{P}^*}[\deg(A)] = \mathbf{d}$  becomes the system of equations (1) from the introduction, and we want to find a solution  $\theta \in \Theta$ .

The system (1) is a special case of a general class that Sanyal, Sturmfels, and Vinzant [31] study using algebraic geometry and matroid theory (extending work of Proudfoot and Speyer [28]). Define

$$\chi(t) = \sum_{k=0}^n \left( \binom{n}{k} + n \binom{n-1}{k} \right) (t-1)_k^{(2)},$$

in which  $\binom{n}{k}$  is the Stirling number of the second kind and  $(x)_{k+1}^{(2)} = x(x-2)\cdots(x-2k)$  is a generalized falling factorial. Then, there is a polynomial  $H(\mathbf{d})$  in the  $d_i$  such that for  $\mathbf{d} \in \mathbb{R}^n$  with  $H(\mathbf{d}) \neq 0$ , the number of solutions  $\theta \in \mathbb{R}^n$  to (1) is  $(-1)^n \chi(0)$ . Moreover, the polynomial  $H(\mathbf{d})$  has degree  $2(-1)^n(n\chi(0) + \chi'(0))$  and characterizes those  $\mathbf{d}$  for which the equations above have multiple roots. We refer to [31] for more details.

Next, we characterize the set of graphic sequences  $\mathcal{W}$  and determine its relation to the mean parameter space  $\mathcal{M}$ . Recall that we say  $\mathbf{d} = (d_1, \dots, d_n)$  is a (weighted) graphic sequence if there is a graph  $G$  with edge weights in  $\mathbb{R}_0$  that realizes  $\mathbf{d}$ . In the case of unweighted graphs we have the combinatorial constraint that there is at most one edge between any pair of vertices, which translates into a set of constraints in the Erdős-Gallai criterion (4). In the case of weighted graphs, on the other hand, every edge can have as much weight as possible, so intuitively we would expect that the criterion for a weighted graphic sequence is simpler than the Erdős-Gallai criterion. This is indeed the case, as the following result shows.

**Lemma 3.4.** *A sequence  $(d_1, \dots, d_n) \in \mathbb{R}_0^n$  is graphic if and only if*

$$\max_{1 \leq i \leq n} d_i \leq \frac{1}{2} \sum_{i=1}^n d_i. \quad (6)$$

*Proof.* Clearly if  $(d_1, \dots, d_n) \in \mathbb{R}_0^n$  is a graphic sequence then so is  $(d_{\pi(1)}, \dots, d_{\pi(n)})$ , for any permutation  $\pi$  of  $\{1, \dots, n\}$ . Thus without loss of generality we can assume  $d_1 \geq d_2 \geq \dots \geq d_n$ , and in this case condition (6) reduces to

$$d_1 \leq \sum_{i=2}^n d_i. \quad (7)$$

It is easy to see that if  $(d_1, \dots, d_n) \in \mathbb{R}_0^n$  is graphic then (7) is satisfied, since the total weight that could come out of vertex 1 is at most  $\sum_{i=2}^n d_i$ . For the converse direction, we first note the following easy properties of weighted graphic sequences:

- (i) The sequence  $(c, c, \dots, c) \in \mathbb{R}_0^n$  is graphic for any  $c \in \mathbb{R}_0$ . For  $n = 2$  this sequence is realized by the graph on 2 vertices having edge weight  $c$ , and for  $n \geq 3$  this sequence is realized by the “chain graph” with weights  $a_{i,i+1} = c/2$  for  $i = 1, \dots, n$  and  $a_{ij} = 0$  otherwise.

- (ii) If  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}_0^n$  satisfies (7) with an equality then  $\mathbf{d}$  is graphic, realized by the “star graph” with weights  $a_{1j} = d_j$  for  $j = 2, \dots, n$  and  $a_{ij} = 0$  otherwise.
- (iii) If  $\mathbf{d} = (d_1, \dots, d_m) \in \mathbb{R}_0^m$  is graphic for some  $m < n$  then so is  $\bar{\mathbf{d}} = (d_1, \dots, d_m, 0, \dots, 0) \in \mathbb{R}_0^n$ . This follows since we can obtain a graph that realizes  $\bar{\mathbf{d}}$  by inserting  $n - m$  isolated vertices to the graph that realizes  $\mathbf{d}$ .
- (iv) If  $\mathbf{d}^{(1)}$  and  $\mathbf{d}^{(2)}$  are graphic then so is  $\mathbf{d}^{(1)} + \mathbf{d}^{(2)}$ . This is because if  $G_1$  and  $G_2$  are graphs that realize  $\mathbf{d}^{(1)}$  and  $\mathbf{d}^{(2)}$ , respectively, then  $\mathbf{d}^{(1)} + \mathbf{d}^{(2)}$  is realized by the graph  $G$  whose edge weights are the sum of the corresponding edge weights in  $G_1$  and  $G_2$ .

Now we prove the converse direction by induction on  $n$ . For the base case  $n = 2$ , condition (7) gives us  $d_1 \leq d_2 \leq d_1$ , so  $(d_1, d_2)$  is graphic by property (i). Assume that the claim holds for  $n - 1$ , and we will prove it also holds for  $n$ . So suppose we have a sequence  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}_0^n$  satisfying (7), and let

$$K = \frac{1}{n-2} \left( \sum_{i=2}^n d_i - d_1 \right) \geq 0$$

If  $K = 0$  then (7) is satisfied with an equality, which implies  $\mathbf{d}$  is graphic by property (ii). So assume  $K > 0$ . We consider two possibilities.

1. Suppose  $K \geq d_n$ . Then we can write  $\mathbf{d} = \mathbf{d}^{(1)} + \mathbf{d}^{(2)}$ , where

$$\mathbf{d}^{(1)} = (d_1 - d_n, d_2 - d_n, \dots, d_{n-1} - d_n, 0) \in \mathbb{R}_0^n$$

and

$$\mathbf{d}^{(2)} = (d_n, d_n, \dots, d_n) \in \mathbb{R}_0^n.$$

The assumption  $K \geq d_n$  implies  $d_1 - d_n \leq \sum_{i=2}^{n-1} (d_i - d_n)$ , so  $(d_1 - d_n, d_2 - d_n, \dots, d_{n-1} - d_n) \in \mathbb{R}_0^{n-1}$  is a graphic sequence by induction hypothesis. This implies  $\mathbf{d}^{(1)}$  is also graphic by property (iii). Furthermore,  $\mathbf{d}^{(2)}$  is graphic by property (i), so  $\mathbf{d} = \mathbf{d}^{(1)} + \mathbf{d}^{(2)}$  is also a graphic sequence by property (iv).

2. Suppose  $K < d_n$ . Then write  $\mathbf{d} = \mathbf{d}^{(3)} + \mathbf{d}^{(4)}$ , where

$$\mathbf{d}^{(3)} = (d_1 - K, d_2 - K, \dots, d_n - K) \in \mathbb{R}_0^n$$

and

$$\mathbf{d}^{(4)} = (K, K, \dots, K) \in \mathbb{R}_0^n.$$

By construction,  $\mathbf{d}^{(3)}$  satisfies  $d_1 - K = \sum_{i=2}^n (d_i - K)$ , so  $\mathbf{d}^{(3)}$  is a graphic sequence by property (ii). Since  $\mathbf{d}^{(4)}$  is also graphic by property (i), we conclude that  $\mathbf{d} = \mathbf{d}^{(3)} + \mathbf{d}^{(4)}$  is graphic by property (iv).

This completes the induction step and finishes the proof.  $\square$

Observe that condition (6) is implied by the case  $k = 1$  in the Erdős-Gallai criterion (4). This means any sequence that satisfies the Erdős-Gallai criterion also satisfies condition (6), which is to be expected, since any unweighted graph is also a weighted graph, so unweighted graphic sequences are also weighted graphic sequences.

Given the criterion above, we can now write the set  $\mathcal{W}$  of graphic sequences explicitly as

$$\mathcal{W} = \left\{ (d_1, \dots, d_n) \in \mathbb{R}_0^n : \max_{1 \leq i \leq n} d_i \leq \frac{1}{2} \sum_{i=1}^n d_i \right\}.$$

We have the following simple property.

**Proposition 3.5.** *The set  $\mathcal{W}$  is convex, and  $\mathcal{M} = \mathcal{W}$ .*

*Proof.* We first prove that  $\mathcal{W}$  is convex. Given  $\mathbf{d} = (d_1, \dots, d_n)$  and  $\mathbf{d}' = (d'_1, \dots, d'_n)$  in  $\mathcal{W}$ , and given  $0 \leq t \leq 1$ , we note that

$$\begin{aligned} \max_{1 \leq i \leq n} (td_i + (1-t)d'_i) &\leq t \max_{1 \leq i \leq n} d_i + (1-t) \max_{1 \leq i \leq n} d'_i \\ &\leq \frac{1}{2}t \sum_{i=1}^n d_i + \frac{1}{2}(1-t) \sum_{i=1}^n d'_i \\ &= \frac{1}{2} \sum_{i=1}^n (td_i + (1-t)d'_i), \end{aligned}$$

which means  $t\mathbf{d} + (1-t)\mathbf{d}' \in \mathcal{W}$ .

Next, recall that we already have  $\mathcal{M} \subseteq \text{conv}(\mathcal{W}) = \mathcal{W}$ , so to conclude  $\mathcal{M} = \mathcal{W}$  it remains to show that  $\mathcal{W} \subseteq \mathcal{M}$ . Given  $\mathbf{d} \in \mathcal{W}$ , let  $G$  be a graph that realizes  $\mathbf{d}$  and let  $\mathbf{w} = (w_{ij})$  be the edge weights of  $G$ , so that  $d_i = \sum_{j \neq i} w_{ij}$  for all  $i = 1, \dots, n$ . Consider a distribution  $\mathbb{P}$  on  $\mathbb{R}_0^{\binom{n}{2}}$  that sets each edge weight  $A_{ij}$  to be an independent exponential random variable with mean parameter  $w_{ij}$ . That is, the density  $p$  of  $\mathbb{P}$  on  $\mathbf{a} = (a_{ij}) \in \mathbb{R}_0^{\binom{n}{2}}$  is

$$p(\mathbf{a}) = \prod_{(i,j)} \frac{1}{w_{ij}} \exp\left(-\frac{a_{ij}}{w_{ij}}\right).$$

Then clearly  $\mathbb{E}_{\mathbb{P}}[A_{ij}] = w_{ij}$ , and thus

$$\mathbb{E}_{\mathbb{P}}[\deg_i(A)] = \sum_{j \neq i} \mathbb{E}_{\mathbb{P}}[A_{ij}] = \sum_{j \neq i} w_{ij} = d_i \quad \text{for } i = 1, \dots, n.$$

This shows that  $\mathbf{d} \in \mathcal{M}$ , as desired.  $\square$

Noting that the mean function  $\mu(t) = 1/t$  is strictly decreasing on  $\text{Dom}(Z_1) = \mathbb{R}_+$ , we reach the following conclusion.

**Corollary 3.6.** *Given  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}^n$ , the system of MLE equations*

$$d_i = \sum_{j \neq i} \frac{1}{\theta_i + \theta_j} \quad \text{for } i = 1, \dots, n \tag{8}$$

*has a valid solution  $\theta = (\theta_1, \dots, \theta_n) \in \Theta$  if and only if*

$$\mathbf{d} \in \mathcal{M}^\circ = \left\{ (d'_1, \dots, d'_n) \in \mathbb{R}_+^n : \max_{1 \leq i \leq n} d'_i < \frac{1}{2} \sum_{i=1}^n d'_i \right\}.$$

*Furthermore, if  $\mathbf{d} \in \mathcal{M}^\circ$ , then the valid solution  $\theta \in \Theta$  is unique, and it has the property that*

$$\text{sgn}(d_i - d_j) = \text{sgn}(\theta_j - \theta_i) \quad \text{for all } i \neq j.$$

**Example 3.7.** Let  $n = 3$  and  $\mathbf{d} = (d_1, d_2, d_3) \in \mathbb{R}^n$  with  $d_1 \geq d_2 \geq d_3$ . In this case it is easy to see that the system of equations (8) gives us

$$\frac{1}{\theta_1 + \theta_2} = \frac{1}{2}(d_1 + d_2 - d_3), \quad \frac{1}{\theta_1 + \theta_3} = \frac{1}{2}(d_1 - d_2 + d_3), \quad \text{and} \quad \frac{1}{\theta_2 + \theta_3} = \frac{1}{2}(-d_1 + d_2 + d_3),$$

from which we obtain a unique solution  $\theta = (\theta_1, \theta_2, \theta_3)$ . Recall that  $\theta \in \Theta$  means  $\theta_1 + \theta_2 > 0$ ,  $\theta_1 + \theta_3 > 0$ , and  $\theta_2 + \theta_3 > 0$ , so the equation above tells us that  $\theta \in \Theta$  if and only if  $d_1 < d_2 + d_3$ . In particular, this also implies  $d_3 > d_1 - d_2 \geq 0$ , so  $\mathbf{d} \in \mathbb{R}_+^3$ . Hence there is a unique solution  $\theta \in \Theta$  if and only if  $\mathbf{d} \in \mathcal{M}^\circ$ , as stated in Corollary 3.6.

### 3.3 Weighted graphs with discrete weights

Finally, let  $S = \mathbb{N}_0$  and  $\nu$  be the counting measure, so we are considering weighted graphs with discrete (and unbounded) weights. In this case

$$Z_1(t) = \log \sum_{a=0}^{\infty} \exp(-ta) = \begin{cases} -\log(1 - \exp(-t)) & \text{if } t > 0 \\ \infty & \text{if } t \leq 0. \end{cases}$$

Thus  $\text{Dom}(Z_1) = (0, \infty)$ , and the natural parameter space is

$$\Theta = \{(\theta_1, \dots, \theta_n) \in \mathbb{R}^n : \theta_i + \theta_j > 0 \text{ for } i \neq j\},$$

which is open. Given  $\theta \in \Theta$ , the edge weights  $A_{ij}$  are independent geometric random variables with probability mass function

$$\mathbb{P}^*(A_{ij} = a) = (1 - \exp(-\theta_i - \theta_j)) \exp(-(\theta_i + \theta_j)a) \quad \text{for } a \in \mathbb{N}_0.$$

The mean parameters are now

$$\mathbb{E}_{\mathbb{P}^*}[A_{ij}] = \frac{\exp(-\theta_i - \theta_j)}{1 - \exp(-\theta_i - \theta_j)} = \frac{1}{\exp(\theta_i + \theta_j) - 1},$$

so the MLE equation  $\mathbb{E}_{\mathbb{P}^*}[\deg(A)] = \mathbf{d}$  now becomes the system of equations

$$d_i = \sum_{j \neq i} \frac{1}{\exp(\theta_i + \theta_j) - 1} \quad \text{for } i = 1, \dots, n,$$

and we want to find a solution  $\theta \in \Theta$ .

Since  $\nu^{\binom{n}{2}}$  is the counting measure on  $\mathbb{N}_0^{\binom{n}{2}}$ ,  $\mathfrak{P}$  contains all the Dirac measures, so we have  $\mathcal{M} = \text{conv}(\mathcal{W})$  from the general discussion in Section 2. We now characterize  $\mathcal{W}$ , the set of all graphic sequences. As in Lemma 3.4, we also have a simple characterization for when  $\mathbf{d} = (d_1, \dots, d_n)$  is a degree sequence of a graph  $G$  with edge weights in  $\mathbb{N}_0$ . The proof of the following result is inspired by [9].

**Lemma 3.8.** *A sequence  $(d_1, \dots, d_n) \in \mathbb{N}_0^n$  is graphic if and only if  $\sum_{i=1}^n d_i$  is even and*

$$\max_{1 \leq i \leq n} d_i \leq \frac{1}{2} \sum_{i=1}^n d_i.$$

*Proof.* Without loss of generality we may assume  $d_1 \geq d_2 \geq \dots \geq d_n$ , so the inequality above becomes  $d_1 \leq \sum_{i=2}^n d_i$ . The necessary part is easy: if  $(d_1, \dots, d_n)$  is a degree sequence of a graph  $G$  with edge weights  $a_{ij} \in \mathbb{N}_0$ , then  $\sum_{i=1}^n d_i = 2 \sum_{(i,j)} a_{ij}$  is even, and the total weight coming out of vertex 1 is at most  $\sum_{i=2}^n d_i$ . The converse direction is trivial if  $n = 2$ , so assume  $n \geq 3$ . We proceed by induction on  $s = \sum_{i=1}^n d_i$ . The statement is clearly true for  $s = 0$  and  $s = 2$ . Assume the statement is true for some even  $s \in \mathbb{N}$ , and suppose we are given  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{N}_0^n$  with  $d_1 \geq \dots \geq d_n$ ,  $\sum_{i=1}^n d_i = s + 2$ , and  $d_1 \leq \sum_{i=2}^n d_i$ . Without loss of generality we may assume  $d_n \geq 1$ , for otherwise we can proceed with only the nonzero elements of  $\mathbf{d}$ . Let  $k$  be the smallest index such that  $d_k > d_{k+1}$ , with  $k = n - 1$  if  $d_1 = \dots = d_n$ , and let  $\mathbf{d}' = (d_1, \dots, d_{k-1}, d_k - 1, d_{k+1}, \dots, d_n - 1)$ . We will show that  $\mathbf{d}'$  is graphic. This will imply that  $\mathbf{d}$  is graphic, because if  $G'$  is a graph with edge weights  $a'_{ij}$  that realizes  $\mathbf{d}'$ , then  $\mathbf{d}$  is realized by the graph  $G$  with edge weights  $a_{kn} = a'_{kn} + 1$  and  $a_{ij} = a'_{ij}$  otherwise.

Now for  $\mathbf{d}' = (d'_1, \dots, d'_n)$  given above, we have  $d'_1 \geq \dots \geq d'_n$  and  $\sum_{i=1}^n d'_i = \sum_{i=1}^n d_i - 2 = s$  is even. So it suffices to show that  $d'_1 \leq \sum_{i=2}^n d'_i$ , for then we can apply the induction hypothesis to conclude that  $\mathbf{d}'$  is graphic. If  $k = 1$ , then  $d'_1 = d_1 - 1 \leq \sum_{i=2}^n d_i - 1 = \sum_{i=2}^n d'_i$ . If  $k > 1$  then  $d_1 = d_2$ , so  $d_1 < \sum_{i=2}^n d_i$  since  $d_n \geq 1$ . In particular, since  $\sum_{i=1}^n d_i$  is even,  $\sum_{i=2}^n d_i - d_1 = \sum_{i=1}^n d_i - 2d_1$  is also even, hence  $\sum_{i=2}^n d_i - d_1 \geq 2$ . Therefore,  $d'_1 = d_1 \leq \sum_{i=2}^n d_i - 2 = \sum_{i=2}^n d'_i$ . This finishes the proof of the lemma.  $\square$

The criterion above allows us to write an explicit form for  $\mathcal{W}$ ,

$$\mathcal{W} = \left\{ (d_1, \dots, d_n) \in \mathbb{N}_0^n : \sum_{i=1}^n d_i \text{ is even and } \max_{1 \leq i \leq n} d_i \leq \frac{1}{2} \sum_{i=1}^n d_i \right\}.$$

Now we need to characterize  $\text{conv}(\mathcal{W})$ . Let  $\mathcal{W}_1$  denote the set of all graphic sequences from Section 3.2, when the edge weights are in  $\mathbb{R}_0$ ,

$$\mathcal{W}_1 = \left\{ (d_1, \dots, d_n) \in \mathbb{R}_0^n : \max_{1 \leq i \leq n} d_i \leq \frac{1}{2} \sum_{i=1}^n d_i \right\}.$$

It turns out that when we take the convex hull of  $\mathcal{W}$ , we essentially recover  $\mathcal{W}_1$ .

**Lemma 3.9.**  $\overline{\text{conv}(\mathcal{W})} = \mathcal{W}_1$ .

*Proof.* Clearly  $\mathcal{W} \subseteq \mathcal{W}_1$ , so  $\overline{\text{conv}(\mathcal{W})} \subseteq \mathcal{W}_1$  since  $\mathcal{W}_1$  is closed and convex, by Proposition 3.5. Conversely, let  $\mathbb{Q}$  denote the set of rational numbers. We will first show that  $\mathcal{W}_1 \cap \mathbb{Q}^n \subseteq \text{conv}(\mathcal{W})$  and then proceed by a limit argument. Let  $\mathbf{d} \in \mathcal{W}_1 \cap \mathbb{Q}^n$ , so  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{Q}^n$  with  $d_i \geq 0$  and  $\max_{1 \leq i \leq n} d_i \leq \frac{1}{2} \sum_{i=1}^n d_i$ . Choose  $K \in \mathbb{N}$  large enough such that  $Kd_i \in \mathbb{N}_0$  for all  $i = 1, \dots, n$ . Observe that  $2K\mathbf{d} = (2Kd_1, \dots, 2Kd_n) \in \mathbb{N}_0^n$  has the property that  $\sum_{i=1}^n 2Kd_i \in \mathbb{N}_0$  is even and  $\max_{1 \leq i \leq n} 2Kd_i \leq \frac{1}{2} \sum_{i=1}^n 2Kd_i$ , so  $2K\mathbf{d} \in \mathcal{W}$  by definition. Since  $0 = (0, \dots, 0) \in \mathcal{W}$  as well, all elements along the segment joining 0 and  $2K\mathbf{d}$  lie in  $\text{conv}(\mathcal{W})$ , so in particular,  $\mathbf{d} = (2K\mathbf{d})/(2K) \in \text{conv}(\mathcal{W})$ . This shows that  $\mathcal{W}_1 \cap \mathbb{Q}^n \subseteq \text{conv}(\mathcal{W})$ , and hence  $\overline{\mathcal{W}_1 \cap \mathbb{Q}^n} \subseteq \overline{\text{conv}(\mathcal{W})}$ .

To finish the proof it remains to show that  $\overline{\mathcal{W}_1 \cap \mathbb{Q}^n} = \mathcal{W}_1$ . On the one hand we have

$$\overline{\mathcal{W}_1 \cap \mathbb{Q}^n} \subseteq \overline{\mathcal{W}_1} \cap \overline{\mathbb{Q}^n} = \mathcal{W}_1 \cap \mathbb{R}_0^n = \mathcal{W}_1.$$

For the other direction, given  $\mathbf{d} \in \mathcal{W}_1$ , choose  $\mathbf{d}_1, \dots, \mathbf{d}_n \in \mathcal{W}_1$  such that  $\mathbf{d}, \mathbf{d}_1, \dots, \mathbf{d}_n$  are in general position, so that the convex hull  $C$  of  $\mathbf{d}, \mathbf{d}_1, \dots, \mathbf{d}_n$  is full dimensional. This can be done, for instance, by noting that the following  $n+1$  points in  $\mathcal{W}_1$  are in general position:

$$0, \mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_3, \dots, \mathbf{e}_1 + \mathbf{e}_n, \mathbf{e}_1 + \mathbf{e}_2 + \dots + \mathbf{e}_n$$

where  $\mathbf{e}_1, \dots, \mathbf{e}_n$  are the standard basis of  $\mathbb{R}^n$ . For each  $m \in \mathbb{N}$  and  $i = 1, \dots, n$ , choose  $\mathbf{d}_i^{(m)}$  on the line segment between  $\mathbf{d}$  and  $\mathbf{d}_i$  such that the convex hull  $C_m$  of  $\mathbf{d}, \mathbf{d}_1^{(m)}, \dots, \mathbf{d}_n^{(m)}$  is full dimensional and has diameter at most  $1/m$ . Since  $C_m$  is full dimensional we can choose a rational point  $\mathbf{r}_m \in C_m \subseteq C \subseteq \mathcal{W}_1$ . Thus we have constructed a sequence of rational points  $(\mathbf{r}_m)$  in  $\mathcal{W}_1$  converging to  $\mathbf{d}$ , which shows that  $\mathcal{W}_1 \subseteq \overline{\mathcal{W}_1 \cap \mathbb{Q}^n}$ .  $\square$

**Remark 3.10.** Since  $\mathcal{W}$  is countable and discrete, i.e.  $\|\mathbf{d} - \mathbf{d}'\|_\infty \geq 1$  for  $\mathbf{d}, \mathbf{d}' \in \mathcal{W}$  with  $\mathbf{d} \neq \mathbf{d}'$ , it seems that  $\text{conv}(\mathcal{W})$  is closed, so we in fact have  $\text{conv}(\mathcal{W}) = \mathcal{W}_1$ . However, since we are only interested in the interior of  $\text{conv}(\mathcal{W})$ , knowing  $\overline{\text{conv}(\mathcal{W})} = \mathcal{W}_1$  is enough for our purposes.

Recalling that a convex set and its closure have the same interior points, the result above gives us

$$\mathcal{M}^\circ = \text{conv}(\mathcal{W})^\circ = (\overline{\text{conv}(\mathcal{W})})^\circ = \mathcal{W}_1^\circ = \left\{ (d_1, \dots, d_n) \in \mathbb{R}_+^n : \max_{1 \leq i \leq n} d_i < \frac{1}{2} \sum_{i=1}^n d_i \right\}.$$

Furthermore, noting that the mean function  $\mu(t) = 1/(\exp(t) - 1)$  is strictly decreasing for  $t \in \text{Dom}(Z_1) = (0, \infty)$ , we conclude the following.

**Corollary 3.11.** *Given  $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}^n$ , the system of MLE equations*

$$d_i = \sum_{j \neq i} \frac{1}{\exp(\theta_i + \theta_j) - 1} \quad \text{for } i = 1, \dots, n \quad (9)$$

has a valid solution  $\theta = (\theta_1, \dots, \theta_n) \in \Theta$  if and only if

$$\mathbf{d} \in \mathcal{M}^\circ = \left\{ (d'_1, \dots, d'_n) \in \mathbb{R}_+^n : \max_{1 \leq i \leq n} d'_i < \frac{1}{2} \sum_{i=1}^n d'_i \right\}.$$

Furthermore, if  $\mathbf{d} \in \mathcal{M}^\circ$ , then the valid solution  $\theta \in \Theta$  is unique, and it has the property that

$$\text{sgn}(d_i - d_j) = \text{sgn}(\theta_j - \theta_i) \quad \text{for all } i \neq j.$$

**Example 3.12.** Let  $n = 3$  and  $\mathbf{d} = (d_1, d_2, d_3) \in \mathbb{R}^n$  with  $d_1 \geq d_2 \geq d_3$ . One can easily check that the system of equations (9) gives us

$$\begin{aligned} \theta_1 + \theta_2 &= \log \left( 1 + \frac{2}{d_1 + d_2 - d_3} \right) \\ \theta_1 + \theta_3 &= \log \left( 1 + \frac{2}{d_1 - d_2 + d_3} \right) \\ \theta_2 + \theta_3 &= \log \left( 1 + \frac{2}{-d_1 + d_2 + d_3} \right), \end{aligned}$$

from which we can obtain  $\theta = (\theta_1, \theta_2, \theta_3)$ . For  $\theta$  to be in  $\Theta$  we want  $\theta_1 + \theta_2 > 0$ ,  $\theta_1 + \theta_3 > 0$ , and  $\theta_2 + \theta_3 > 0$ , which means  $2/(-d_1 + d_2 + d_3) > 0$ , or equivalently,  $d_1 < d_2 + d_3$ . This also implies  $d_3 > d_1 - d_2 \geq 0$ , so  $\mathbf{d} \in \mathbb{R}_+^3$ . Thus  $\theta \in \Theta$  if and only if  $\mathbf{d} \in \mathcal{M}^\circ$ , as stated in Corollary 3.11.

## 4 Existence and consistency of the MLE from one sample

In this section we study the existence and consistency of the MLE of the parameters from one graph sample. Given  $\theta \in \Theta$ , let  $\hat{A} = (\hat{A}_{ij})$  be a sample drawn from the random graph distribution parameterized by  $\theta$ . Let  $\hat{\mathbf{d}}$  be the empirical degree sequence associated with  $\hat{A}$ . As we saw in Section 2, the MLE  $\hat{\theta} \in \Theta$  of  $\theta$  is the solution to the system of equations (3). The MLE  $\hat{\theta}$  is *consistent* if  $\hat{\theta}$  converges in probability to  $\theta$  as  $n \rightarrow \infty$ .

In the case of unweighted graphs, Chatterjee, Diaconis, and Sly [8] showed that with high probability the MLE  $\hat{\theta}$  exists and is consistent.

**Theorem 4.1** ([8, Theorem 1.3]). *In the unweighted graph case, let  $M = \|\theta\|_\infty$ . Then with probability at least  $1 - C(M)/n^2$  the MLE  $\hat{\theta}$  exists and satisfies*

$$\|\hat{\theta} - \theta\|_\infty \leq C(M) \sqrt{\frac{\log n}{n}},$$

where  $C(M)$  is a constant depending on  $M$ .

For the case of weighted graphs, recall that  $\theta \in \Theta$  means  $\theta_i + \theta_j > 0$  for all  $i \neq j$ . Our goal in this section is to prove the following consistency results.

**Theorem 4.2.** *In the case of weighted graphs with continuous weights, let  $L = \min_{i \neq j} \theta_i + \theta_j$  and  $M = \max_{i \neq j} \theta_i + \theta_j$ , so  $0 < L \leq M$ . Let  $k > 1$  be fixed. Then for sufficiently large  $n$ , with probability at least  $1 - 3n^{-(k-1)}$  the MLE  $\hat{\theta}$  exists and satisfies*

$$\|\hat{\theta} - \theta\|_\infty \leq \frac{150M^2}{L} \sqrt{\frac{k \log n}{n}}.$$

**Theorem 4.3.** *In the case of weighted graphs with discrete weights, let  $L = \min_{i \neq j} \theta_i + \theta_j$  and  $M = \max_{i \neq j} \theta_i + \theta_j$ , so  $0 < L \leq M$ . Let  $k > 1$  be fixed. Then for sufficiently large  $n$ , with probability at least  $1 - c(M)^n - 3n^{-(k-1)}$  the MLE  $\hat{\theta}$  exists and satisfies*

$$\|\hat{\theta} - \theta\|_\infty \leq \frac{(\exp(5M) - 1)^2}{\exp(5M)} \sqrt{\frac{12}{\exp(L/2) - 1}} \sqrt{\frac{k \log n}{n}},$$

where  $0 < c(M) < 1$  is a constant that depends on  $M$ .

#### 4.1 Proof of Theorem 4.2

We are now working with weighted graphs with continuous weights. Recall that in this case the edge weights  $A_{ij}$  are exponential random variables with mean  $\mu(\theta_i + \theta_j) = 1/(\theta_i + \theta_j)$ .

For the existence of the MLE  $\hat{\theta}$ , recall from Proposition 2.1 that  $\hat{\theta}$  exists if and only if the empirical degree sequence  $\hat{\mathbf{d}}$  belongs to the interior of the mean parameter space  $\mathcal{M}^\circ$ . Since  $\hat{\mathbf{d}}$  necessarily belongs to  $\mathcal{M}$ , the MLE  $\hat{\theta}$  does not exist precisely when  $\hat{\mathbf{d}}$  falls on the boundary  $\partial\mathcal{M} = \mathcal{M} \setminus \mathcal{M}^\circ$ . From Corollary 3.6 we see that this boundary is given by

$$\partial\mathcal{M} = \left\{ \mathbf{d}' \in \mathbb{R}_0^n : d'_i = 0 \text{ for some } i \text{ or } \max_{1 \leq i \leq n} d'_i = \frac{1}{2} \sum_{i=1}^n d'_i \right\},$$

which has Lebesgue measure 0. Since the distribution on  $A$  is continuous and  $\mathbf{d}$  is a continuous function of  $A$ , we have  $\mathbf{d} \in \partial\mathcal{M}$  with probability 0. Thus, in this case the MLE  $\hat{\theta}$  exists almost surely.

Now for the consistency of  $\hat{\theta}$ , recall from Section 2 that the MLE  $\hat{\theta}$  satisfies the equation  $-\nabla Z(\hat{\theta}) = \hat{\mathbf{d}}$ . Let  $\mathbf{d} = -\nabla Z(\theta)$  denote the expected degree sequence of the maximum entropy distribution with parameter  $\theta$ . By the mean value theorem for vector-valued functions [20, p. 341], we can write

$$\mathbf{d} - \hat{\mathbf{d}} = \nabla Z(\hat{\theta}) - \nabla Z(\theta) = J(\hat{\theta} - \theta). \quad (10)$$

Here  $J$  is a matrix obtained by integrating (element-wise) the Hessian of  $Z$  on intermediate points between  $\theta$  and  $\hat{\theta}$ :

$$J = \int_0^1 \nabla^2 Z(t\theta + (1-t)\hat{\theta}) dt.$$

At any point  $\xi = t\theta + (1-t)\hat{\theta}$ , the gradient  $\nabla Z(\xi)$  is

$$(\nabla Z(\xi))_i = -\sum_{j \neq i} \mu(\xi_i + \xi_j) = -\sum_{j \neq i} \frac{1}{\xi_i + \xi_j}.$$

Therefore, the Hessian is given by

$$(\nabla^2 Z(\xi))_{ij} = \frac{1}{(\xi_i + \xi_j)^2} \text{ for } i \neq j \quad \text{and} \quad (\nabla^2 Z(\xi))_{ii} = \sum_{j \neq i} \frac{1}{(\xi_i + \xi_j)^2} = \sum_{j \neq i} (\nabla^2 Z(\xi))_{ij}.$$

We call a nonnegative matrix such as  $\nabla^2 Z$  *diagonally balanced* if each diagonal entry is equal to the sum of the other entries in its row. Since  $\theta, \theta' \in \Theta$  and we assume  $\theta_i + \theta_j \leq M$ , it follows that for  $i \neq j$ ,

$$0 < \xi_i + \xi_j \leq \max\{\theta_i + \theta_j, \hat{\theta}_i + \hat{\theta}_j\} \leq \max\{M, 2\|\hat{\theta}\|_\infty\} \leq M + 2\|\hat{\theta}\|_\infty.$$

Thus, the off-diagonal entries of  $\nabla^2 Z(\xi)$  are bounded below by  $1/(M + 2\|\hat{\theta}\|_\infty)^2$ , and so  $J$  is a symmetric and diagonally balanced matrix with off-diagonal entries bounded below by  $1/(M + 2\|\hat{\theta}\|_\infty)^2$ , being an average of such matrices. By the main result of [15],  $J$  is invertible and its inverse satisfies

$$\|J^{-1}\|_\infty \leq \frac{(M + 2\|\hat{\theta}\|_\infty)^2(3n - 4)}{2(n - 1)(n - 2)} \leq \frac{2}{n} (M + 2\|\hat{\theta}\|_\infty)^2,$$

where the last inequality holds for sufficiently large  $n$ . Inverting  $J$  in (10) and applying the bound above gives

$$\|\theta - \hat{\theta}\|_\infty \leq \|J^{-1}\|_\infty \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty \leq \frac{2}{n} (M + 2\|\hat{\theta}\|_\infty)^2 \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty. \quad (11)$$

Since  $A_{ij}$  is an exponential random variable with rate  $\lambda = \theta_i + \theta_j \geq L$ , Lemma A.2 tells us that  $A_{ij} - 1/(\theta_i + \theta_j)$  is a  $(4/L^2, L/2)$ -subexponential random variable. Moreover, since  $(A_{ij}, i \neq j)$  are independent, we can apply the concentration inequality for subexponential random variables [14]. Given any constant  $k > 1$ , for sufficiently large  $n$  and for each  $i = 1, \dots, n$ , we have

$$\begin{aligned} \mathbb{P}\left(|\hat{d}_i - d_i| \geq \sqrt{\frac{9kn \log n}{L^2}}\right) &\leq \mathbb{P}\left(|\hat{d}_i - d_i| \geq \sqrt{\frac{8k(n-1) \log(n-1)}{L^2}}\right) \\ &= \mathbb{P}\left(\left|\frac{1}{n-1} \sum_{j \neq i} \left(A_{ij} - \frac{1}{\theta_i + \theta_j}\right)\right| \geq \sqrt{\frac{8k \log(n-1)}{L^2(n-1)}}\right) \\ &\leq 2 \exp\left(-\frac{L^2(n-1)}{8} \frac{8k \log(n-1)}{L^2(n-1)}\right) \\ &= \frac{2}{(n-1)^k} \leq \frac{3}{n^k}, \end{aligned}$$

and so by the union bound,

$$\mathbb{P}\left(\|\mathbf{d} - \hat{\mathbf{d}}\|_\infty \geq \sqrt{\frac{9kn \log n}{L^2}}\right) = \mathbb{P}\left(|\hat{d}_i - d_i| \geq \sqrt{\frac{9kn \log n}{L^2}} \text{ for } i = 1, \dots, n\right) \leq \frac{3}{n^{k-1}}.$$

Assume now that  $\|\mathbf{d} - \hat{\mathbf{d}}\|_\infty \leq \sqrt{9kn \log n / L^2}$ , which happens with probability at least  $1 - 3n^{-(k-1)}$ . Then from (11) and using the triangle inequality, we get

$$\|\hat{\theta}\|_\infty \leq \|\theta - \hat{\theta}\|_\infty + \|\theta\|_\infty \leq \frac{6}{L} \sqrt{\frac{k \log n}{n}} (M + 2\|\hat{\theta}\|_\infty)^2 + M. \quad (12)$$

What we have shown is that  $\|\hat{\theta}\|_\infty$  satisfies the inequality  $G_n(\|\hat{\theta}\|_\infty) \geq 0$ , where  $G_n(x)$  is the quadratic function

$$G_n(x) = \frac{6}{L} \sqrt{\frac{k \log n}{n}} (M + 2x)^2 - x + M.$$

It is easy to check that for sufficiently large  $n$  we have  $G_n(2M) < 0$  and  $G_n(\log n) < 0$ . Thus,  $G_n(\|\hat{\theta}\|_\infty) \geq 0$  means either  $\|\hat{\theta}\|_\infty < 2M$  or  $\|\hat{\theta}\|_\infty > \log n$ . We claim that for sufficiently large  $n$ ,  $\|\hat{\theta}\|_\infty < 2M$ . Suppose the contrary that  $\|\hat{\theta}\|_\infty > \log n$ . Since  $\hat{\theta}_i + \hat{\theta}_j > 0$  for each  $i \neq j$ , there can be at most one index  $i$  with  $\hat{\theta}_i < 0$ . We consider two cases:

1. **Case 1:**  $\hat{\theta}_i \geq 0$  for all  $i = 1, \dots, n$ . Let  $i^*$  be an index with  $\hat{\theta}_{i^*} = \|\hat{\theta}\|_\infty > \log n$ . Then, since  $\hat{\theta}_{i^*} + \hat{\theta}_j \geq \hat{\theta}_{i^*}$  for  $j \neq i^*$ ,

$$\begin{aligned} \frac{1}{M} &\leq \frac{1}{n-1} \sum_{j \neq i^*} \frac{1}{\theta_{i^*} + \theta_j} \\ &\leq \frac{1}{n-1} \left| \sum_{j \neq i^*} \frac{1}{\theta_{i^*} + \theta_j} - \sum_{j \neq i^*} \frac{1}{\hat{\theta}_{i^*} + \hat{\theta}_j} \right| + \frac{1}{n-1} \sum_{j \neq i^*} \frac{1}{\hat{\theta}_{i^*} + \hat{\theta}_j} \\ &\leq \frac{1}{n-1} \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty + \frac{1}{\|\hat{\theta}\|_\infty} \\ &\leq \frac{3\sqrt{kn \log n}}{L(n-1)} + \frac{1}{\log n}, \end{aligned}$$

which cannot hold for sufficiently large  $n$ , as the right hand side on the last line tends to 0.

2. **Case 2:**  $\hat{\theta}_i < 0$  for some  $i = 1, \dots, n$ . Without loss of generality assume  $\hat{\theta}_1 < 0 < \hat{\theta}_2 \leq \dots \leq \hat{\theta}_n$ . Following the same chain of inequalities as in the previous case, we obtain

$$\begin{aligned} \frac{1}{M} &\leq \frac{1}{n-1} \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty + \frac{1}{n-1} \left( \frac{1}{\hat{\theta}_n + \hat{\theta}_1} + \sum_{j=2}^{n-1} \frac{1}{\hat{\theta}_j + \hat{\theta}_n} \right) \\ &\leq \frac{3\sqrt{kn \log n}}{L(n-1)} + \frac{1}{(n-1)(\hat{\theta}_n + \hat{\theta}_1)} + \frac{n-2}{(n-1)\|\hat{\theta}\|_\infty} \\ &\leq \frac{3\sqrt{kn \log n}}{L(n-1)} + \frac{1}{(n-1)(\hat{\theta}_n + \hat{\theta}_1)} + \frac{1}{\log n}, \end{aligned}$$

so for sufficiently large  $n$ ,

$$\frac{1}{\hat{\theta}_1 + \hat{\theta}_n} \geq (n-1) \left( \frac{1}{M} - \frac{3\sqrt{kn \log n}}{L(n-1)} - \frac{1}{\log n} \right) \geq \frac{n}{2M},$$

and thus  $\hat{\theta}_1 + \hat{\theta}_i \leq \hat{\theta}_1 + \hat{\theta}_n \leq 2M/n$  for each  $i = 2, \dots, n$ . However, then

$$\begin{aligned} \frac{3\sqrt{kn \log n}}{L} &\geq \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty \geq -\sum_{j=2}^n \frac{1}{\theta_1 + \theta_j} + \sum_{j=2}^n \frac{1}{\hat{\theta}_1 + \hat{\theta}_n} \\ &\geq -\frac{(n-1)}{M} + \frac{n(n-1)}{2M} = \frac{(n-1)(n-2)}{2M}, \end{aligned}$$

which cannot hold for sufficiently large  $n$ .

The analysis above shows that  $\|\hat{\theta}\|_\infty < 2M$  for all sufficiently large  $n$ . Finally, from (11) we conclude that for sufficiently large  $n$ , with probability at least  $1 - 3n^{-(k-1)}$  we have

$$\|\theta - \hat{\theta}\|_\infty \leq \frac{2}{n} (5M)^2 \frac{3\sqrt{kn \log n}}{L} = \frac{150M^2}{L} \sqrt{\frac{k \log n}{n}},$$

as desired.

## 4.2 Proof of Theorem 4.3

We are now working with weighted graphs with discrete weights. Recall that in this case the edge weights  $A_{ij}$  are geometric random variables with mean  $\mu(\theta_i + \theta_j) = 1/(\exp(\theta_i + \theta_j) - 1)$ .

As in the proof of Theorem 4.2, the MLE  $\hat{\theta}$  does not exist if and only if the empirical degree sequence  $\mathbf{d}$  falls on the boundary  $\partial\mathcal{M}$ , which by Corollary 3.11 is given by

$$\partial\mathcal{M} = \left\{ \mathbf{d}' \in \mathbb{R}_0^n : d'_i = 0 \text{ for some } i \text{ or } \max_{1 \leq i \leq n} d'_i = \frac{1}{2} \sum_{i=1}^n d'_i \right\}.$$

Using union bound and the fact that the edge weights  $A_{ij}$  are independent, we have

$$\begin{aligned} \mathbb{P}(d_i = 0 \text{ for some } i) &\leq \sum_{i=1}^n \mathbb{P}(d_i = 0) = \sum_{i=1}^n \mathbb{P}(A_{ij} = 0 \text{ for all } j \neq i) \\ &= \sum_{i=1}^n \prod_{j \neq i} (1 - \exp(-\theta_i - \theta_j)) \leq n (1 - \exp(-M))^{n-1}. \end{aligned}$$

Furthermore, again by union bound,

$$\mathbb{P}\left(\max_{1 \leq i \leq n} d_i = \frac{1}{2} \sum_{i=1}^n d_i\right) = \mathbb{P}\left(d_i = \sum_{j \neq i} d_j \text{ for some } i\right) \leq \sum_{i=1}^n \mathbb{P}\left(d_i = \sum_{j \neq i} d_j\right).$$

Note that we have  $d_i = \sum_{j \neq i} d_j$  for some  $i$  if and only if the edge weights  $A_{jk} = 0$  for all  $j, k \neq i$ . This occurs with the probability

$$\mathbb{P}(A_{jk} = 0 \text{ for } j, k \neq i) = \prod_{\substack{j, k \neq i \\ j \neq k}} (1 - \exp(-\theta_j - \theta_k)) = (1 - \exp(-M))^{\binom{n-1}{2}}.$$

Therefore,

$$\begin{aligned} \mathbb{P}(\mathbf{d} \in \partial \mathcal{M}) &\leq \mathbb{P}(d_i = 0 \text{ for some } i) + \mathbb{P}\left(\max_{1 \leq i \leq n} d_i = \frac{1}{2} \sum_{i=1}^n d_i\right) \\ &\leq n(1 - \exp(-M))^{n-1} + n(1 - \exp(-M))^{\binom{n-1}{2}} \\ &\leq c(M)^n \end{aligned}$$

for sufficiently large  $n$ , where  $0 < c(M) < 1$  is a constant depending on  $M$ . This shows that the MLE  $\hat{\theta}$  exists with probability at least  $1 - c(M)^n$  for sufficiently large  $n$ .

Now assume that the MLE  $\hat{\theta}$  exists. The proof of the consistency of  $\hat{\theta}$  follows the same outline as in the proof of Theorem 4.2. Let  $\mathbf{d} = -\nabla Z(\theta)$  denote the expected degree sequence of the distribution with parameter  $\theta$ . By the mean value theorem, we can write

$$\mathbf{d} - \hat{\mathbf{d}} = \nabla Z(\hat{\theta}) - \nabla Z(\theta) = J(\hat{\theta} - \theta), \quad (13)$$

where  $J = \int_0^1 \nabla^2 Z(t\theta + (1-t)\hat{\theta}) dt$ . In this case, at any point  $\xi = t\theta + (1-t)\hat{\theta}$  the gradient  $\nabla Z(\theta)$  is

$$(\nabla Z(\xi))_i = - \sum_{j \neq i} \frac{1}{\exp(\xi_i + \xi_j) - 1},$$

and the Hessian is given by

$$(\nabla^2 Z(\xi))_{ij} = \frac{\exp(\xi_i + \xi_j)}{(\exp(\xi_i + \xi_j) - 1)^2} \text{ for } i \neq j \quad \text{and} \quad (\nabla^2 Z(\xi))_{ii} = \sum_{j \neq i} \frac{\exp(\xi_i + \xi_j)}{(\exp(\xi_i + \xi_j) - 1)^2} = \sum_{j \neq i} (\nabla^2 Z(\xi))_{ij}.$$

Moreover, since  $\theta, \hat{\theta} \in \Theta$  and we assume  $\theta_i + \theta_j \leq M$ , for  $i \neq j$  we have

$$0 < \xi_i + \xi_j \leq \max\{\theta_i + \theta_j, \hat{\theta}_i + \hat{\theta}_j\} \leq \max\{M, 2\|\hat{\theta}\|_\infty\} \leq M + 2\|\hat{\theta}\|_\infty.$$

Therefore,  $J$  is a symmetric and diagonally balanced matrix with off-diagonal entries bounded below by  $\exp(M + 2\|\hat{\theta}\|_\infty)/(\exp(M + 2\|\hat{\theta}\|_\infty) - 1)^2$ . By inverting  $J$  in (13) and applying the bound on  $J^{-1}$  from [15], we obtain

$$\|\theta - \hat{\theta}\|_\infty \leq \|\tilde{J}^{-1}\|_\infty \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty \leq \frac{2}{n} \frac{(\exp(M + 2\|\hat{\theta}\|_\infty) - 1)^2}{\exp(M + 2\|\hat{\theta}\|_\infty)} \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty. \quad (14)$$

Since  $A_{ij}$  is a geometric random variable with emission probability

$$p = 1 - \exp(-\theta_i - \theta_j) \geq 1 - \exp(-L),$$

Lemma A.3 tells us that  $A_{ij} - 1/(\exp(\theta_i + \theta_j) - 1)$  is  $(\sigma^2, d)$ -subexponential with

$$\sigma^2 = \frac{\sqrt{1-p}}{(1-\sqrt{1-p})^2} \leq \frac{1}{\exp(L/2) - 1} \quad \text{and} \quad d = -\frac{1}{2} \log(1-p) \geq -\frac{1}{2} \log \exp(-L) = \frac{L}{2}.$$

Furthermore, since the edge weights  $(A_{ij}, i \neq j)$  are independent, we can apply the concentration inequality for subexponential random variables [14]. Given any constant  $k > 1$ , for sufficiently large  $n$  and for each  $i = 1, \dots, n$  we have

$$\begin{aligned} \mathbb{P} \left( |\hat{d}_i - d_i| \geq \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} \right) &\leq \mathbb{P} \left( |\hat{d}_i - d_i| \geq \sqrt{\frac{2k(n-1) \log(n-1)}{\exp(L/2) - 1}} \right) \\ &= \mathbb{P} \left( \left| \frac{1}{n-1} \sum_{j \neq i} \left( A_{ij} - \frac{1}{\exp(\theta_i + \theta_j) - 1} \right) \right| \geq \sqrt{\frac{2k \log(n-1)}{(\exp(L/2) - 1)(n-1)}} \right) \\ &\leq 2 \exp \left( -\frac{(\exp(L/2) - 1)(n-1)}{2} \frac{2k \log(n-1)}{(\exp(L/2) - 1)(n-1)} \right) \\ &= \frac{2}{(n-1)^k} \leq \frac{3}{n^k}, \end{aligned}$$

and so by the union bound,

$$\mathbb{P} \left( \|\mathbf{d} - \hat{\mathbf{d}}\|_\infty \geq \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} \right) = \mathbb{P} \left( |\hat{d}_i - d_i| \geq \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} \text{ for } i = 1, \dots, n \right) \leq \frac{3}{n^{k-1}}.$$

Assume now that  $\|\mathbf{d} - \hat{\mathbf{d}}\|_\infty \leq \sqrt{3kn \log n / (\exp(L/2) - 1)}$ , which happens with probability at least  $1 - 3n^{-(k-1)}$ . Then from (14) and using the triangle inequality, we get

$$\|\hat{\theta}\|_\infty \leq \|\theta - \hat{\theta}\|_\infty + \|\theta\|_\infty \leq \sqrt{\frac{12k \log n}{n(\exp(L/2) - 1)}} \frac{(\exp(M + 2\|\hat{\theta}\|_\infty) - 1)^2}{\exp(M + 2\|\hat{\theta}\|_\infty)} + M. \quad (15)$$

This means  $\|\hat{\theta}\|_\infty$  satisfies the inequality  $H_n(\|\hat{\theta}\|_\infty) \geq 0$ , where  $H_n(x)$  is the function

$$H_n(x) = \sqrt{\frac{12k \log n}{n(\exp(L/2) - 1)}} \frac{(\exp(M + 2x) - 1)^2}{\exp(M + 2x)} - x + M.$$

Note that  $H_n$  is a convex function, so it assumes the value 0 at most twice. Moreover, it is easy to check that for all sufficiently large  $n$ , we have  $H_n(2M) < 0$  and  $H_n(\frac{1}{4} \log n) < 0$ . Therefore,  $H_n(\|\hat{\theta}\|_\infty) \geq 0$  implies either  $\|\hat{\theta}\|_\infty < 2M$  or  $\|\hat{\theta}\|_\infty > \frac{1}{4} \log n$ . We claim that for all sufficiently large  $n$ ,  $\|\hat{\theta}\|_\infty < 2M$ . Suppose the contrary that  $\|\hat{\theta}\|_\infty > \frac{1}{4} \log n$ . Since  $\hat{\theta}_i + \hat{\theta}_j > 0$  for each  $i \neq j$ , there can be at most one index  $i$  with  $\hat{\theta}_i < 0$ . We consider two cases:

1. **Case 1:**  $\hat{\theta}_i \geq 0$  for all  $i = 1, \dots, n$ . Let  $i^*$  be an index with  $\hat{\theta}_{i^*} = \|\hat{\theta}\|_\infty > \frac{1}{4} \log n$ . Then, since

$$\widehat{\theta}_{i^*} + \widehat{\theta}_j \geq \widehat{\theta}_{i^*} \text{ for } j \neq i^*,$$

$$\begin{aligned} \frac{1}{\exp(M) - 1} &\leq \frac{1}{n-1} \sum_{j \neq i^*} \frac{1}{\exp(\theta_{i^*} + \theta_j) - 1} \\ &\leq \frac{1}{n-1} \left| \sum_{j \neq i^*} \frac{1}{\exp(\theta_{i^*} + \theta_j) - 1} - \sum_{j \neq i^*} \frac{1}{\exp(\widehat{\theta}_{i^*} + \widehat{\theta}_j) - 1} \right| + \frac{1}{n-1} \sum_{j \neq i^*} \frac{1}{\exp(\widehat{\theta}_{i^*} + \widehat{\theta}_j) - 1} \\ &\leq \frac{1}{n-1} \|\mathbf{d} - \widehat{\mathbf{d}}\|_\infty + \frac{1}{\exp(\|\widehat{\boldsymbol{\theta}}\|_\infty) - 1} \\ &\leq \frac{1}{(n-1)} \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} + \frac{1}{n^{1/4} - 1}, \end{aligned}$$

which cannot hold for sufficiently large  $n$ , as the right hand side on the last line tends to 0.

2. **Case 2:**  $\widehat{\theta}_i < 0$  for some  $i = 1, \dots, n$ . Without loss of generality assume  $\widehat{\theta}_1 < 0 < \widehat{\theta}_2 \leq \dots \leq \widehat{\theta}_n$ . Following the same chain of inequalities as in the previous case, we obtain

$$\begin{aligned} \frac{1}{\exp(M) - 1} &\leq \frac{1}{n-1} \|\mathbf{d} - \widehat{\mathbf{d}}\|_\infty + \frac{1}{n-1} \left( \frac{1}{\exp(\widehat{\theta}_n + \widehat{\theta}_1) - 1} + \sum_{j=2}^{n-1} \frac{1}{\exp(\widehat{\theta}_j + \widehat{\theta}_n) - 1} \right) \\ &\leq \frac{1}{(n-1)} \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} + \frac{1}{(n-1)(\exp(\widehat{\theta}_n + \widehat{\theta}_1) - 1)} + \frac{n-2}{(n-1)(\exp(\|\widehat{\boldsymbol{\theta}}\|_\infty) - 1)} \\ &\leq \frac{1}{(n-1)} \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} + \frac{1}{(n-1)(\exp(\widehat{\theta}_n + \widehat{\theta}_1) - 1)} + \frac{1}{n^{1/4} - 1}, \end{aligned}$$

so for sufficiently large  $n$ ,

$$\frac{1}{\exp(\widehat{\theta}_1 + \widehat{\theta}_n) - 1} \geq (n-1) \left( \frac{1}{\exp(M) - 1} - \frac{1}{(n-1)} \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} - \frac{1}{n^{1/4} - 1} \right) \geq \frac{n}{2(\exp(M) - 1)}.$$

Therefore, for  $i = 2, \dots, n$  we also have

$$\frac{1}{\exp(\widehat{\theta}_1 + \widehat{\theta}_i) - 1} \geq \frac{1}{\exp(\widehat{\theta}_1 + \widehat{\theta}_n) - 1} \geq \frac{n}{2(\exp(M) - 1)}.$$

However, this implies

$$\begin{aligned} \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} &\geq \|\mathbf{d} - \widehat{\mathbf{d}}\|_\infty \geq - \sum_{j=2}^n \frac{1}{\exp(\theta_1 + \theta_j) - 1} + \sum_{j=2}^n \frac{1}{\exp(\widehat{\theta}_1 + \widehat{\theta}_n) - 1} \\ &\geq - \frac{(n-1)}{\exp(M) - 1} + \frac{n(n-1)}{2(\exp(M) - 1)} = \frac{(n-1)(n-2)}{2(\exp(M) - 1)}, \end{aligned}$$

which cannot hold for sufficiently large  $n$ .

The analysis above shows that  $\|\widehat{\boldsymbol{\theta}}\|_\infty < 2M$  for all sufficiently large  $n$ . Finally, using (14) and taking into account the existence of the MLE, we conclude that for sufficiently large  $n$ , with probability at least  $1 - c(M)^n - 3n^{-(k-1)}$  the MLE  $\widehat{\boldsymbol{\theta}}$  exists and satisfies

$$\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_\infty \leq \frac{2}{n} \frac{(\exp(5M) - 1)^2}{\exp(5M)} \sqrt{\frac{3kn \log n}{\exp(L/2) - 1}} = \frac{(\exp(5M) - 1)^2}{\exp(5M)} \sqrt{\frac{12}{\exp(L/2) - 1}} \sqrt{\frac{k \log n}{n}},$$

as desired.

## A Subexponential random variables

Recall that a zero-mean random variable  $X$  is *subexponential* if  $\mathbb{E}[\exp(tX)] \leq \exp(\sigma^2 t^2/2)$  for all  $|t| \leq d$ , for some parameters  $(\sigma^2, d)$ . In particular, a  $(\sigma^2, d)$ -subexponential random variable  $X$  is also  $(\sigma_+^2, d_-)$ -subexponential for any  $\sigma_+^2 \geq \sigma^2$  and  $d_- \leq d$ . Subexponential random variables satisfy the following concentration inequality [14].

**Theorem A.1.** *Let  $X_1, \dots, X_n$  be independent  $(\sigma^2, d)$ -subexponential random variables. Then*

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i\right| \geq t\right) \leq 2Q_n(t)$$

where

$$Q_n(t) = \begin{cases} \exp(-nt^2/2\sigma^2) & \text{if } 0 < t < d\sigma^2 \\ \exp(-dnt/2) & \text{if } t > d\sigma^2. \end{cases}$$

**Lemma A.2.** *If  $W \sim \text{Exponential}(1/\lambda)$ , then  $Z = W - 1/\lambda$  is  $(4/\lambda^2, \lambda/2)$ -subexponential.*

*Proof.* Recall that  $\mathbb{E}[\exp(tW)] = \lambda/(\lambda - t)$  for  $t < \lambda$ . Consider  $|t| \leq \lambda/2$ , so  $-1/2 \leq t/\lambda \leq 1/2$ . By Taylor expansion,

$$\log\left(1 - \frac{t}{\lambda}\right) + \frac{t}{\lambda} = -\frac{t^2}{\lambda^2} \frac{1}{2(1-\xi)^2} \geq -\frac{t^2}{\lambda^2} \frac{1}{2(1-1/2)^2} = -\frac{2t^2}{\lambda^2}$$

where  $\xi$  is some number between 0 and  $t/\lambda$ . This shows that

$$\mathbb{E}[\exp(tZ)] = \frac{1}{\exp(t/\lambda)(1-t/\lambda)} \leq \exp\left(\frac{2t^2}{\lambda^2}\right) \quad \text{for all } |t| \leq \frac{\lambda}{2},$$

which means  $Z$  is  $(4/\lambda^2, \lambda/2)$ -subexponential. □

**Lemma A.3.** *If  $W \sim \text{Geometric}(p)$ , then  $Z = W - (1-p)/p$  is  $(\sigma^2, d)$ -subexponential with*

$$\sigma^2 = \frac{\sqrt{1-p}}{(1-\sqrt{1-p})^2} \quad \text{and} \quad d = -\frac{1}{2} \log(1-p).$$

*Proof.* Recall that

$$\mathbb{E}[\exp(tW)] = \frac{p}{1 - (1-p)e^t} \quad \text{for } t < -\log(1-p).$$

By Taylor expansion,

$$\log(1 - (1-p)e^t) = \log p - \left(\frac{1-p}{p}\right)t - \frac{t^2}{2} \frac{(1-p)e^\xi}{(1 - (1-p)e^\xi)^2}$$

where  $\xi$  is some number between 0 and  $t$ . Note that for  $|t| \leq -\log(1-p)/2$  we have

$$\frac{(1-p)e^\xi}{(1 - (1-p)e^\xi)^2} \leq \frac{\sqrt{1-p}}{(1 - \sqrt{1-p})^2},$$

so

$$\log\left(\frac{p}{1 - (1-p)e^t}\right) - \left(\frac{1-p}{p}\right)t \leq \frac{t^2}{2} \frac{\sqrt{1-p}}{(1 - \sqrt{1-p})^2} \quad \text{for } |t| \leq -\frac{1}{2} \log(1-p).$$

This shows that

$$\mathbb{E}[\exp(tZ)] = \left(\frac{p}{1 - (1-p)e^t}\right) \exp\left(-\frac{(1-p)t}{p}\right) \leq \exp\left(\frac{t^2 \sqrt{1-p}}{2(1 - \sqrt{1-p})^2}\right) \quad \text{for } |t| \leq -\frac{1}{2} \log(1-p).$$

□

## References

- [1] M. Abeles. *Local Cortical Circuits: An Electrophysiological Study*. Springer, Berlin, 1982.
- [2] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, *A learning algorithm for Boltzmann machines*, Cognitive science, **9** (1985), 147–169.
- [3] W. Bair and C. Koch. *Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey*. Neural computation, 8(6):1185–202, August 1996.
- [4] M. Bethge and P. Berens. *Near-maximum entropy models for binary neural representations of natural images*. Advances in Neural Information Processing Systems, **20**. Cambridge, MA: MIT Press, 2008.
- [5] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. Walczak. *Statistical mechanics for natural flocks of birds*. Proceedings of the National Academy of Sciences, **109** (2012), 4786–4791.
- [6] D. A. Butts, C. Weng, J. Jin, C. I. Yeh, N. A. Lesica, J. M. Alonso, and G. B. Stanley. *Temporal precision in the neural code and the timescales of natural vision*. Nature, 449(7158):92–5, 2007.
- [7] C. E. Carr. *Processing of temporal information in the brain*. Annual Review of Neuroscience, **16** (1993), 223–243.
- [8] S. Chatterjee, P. Diaconis, and A. Sly. *Random graphs with a given degree sequence*. Annals of Applied Probability, **21** (2011), 1400–1435.
- [9] S. A. Choudum. *A simple proof of the Erdős-Gallai theorem on graph sequences*. Bull. Austr. Math. Soc. 33:67–70, 1986.
- [10] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, 2006.
- [11] G. Desbordes, J. Jin, C. Weng, N. A. Lesica, G. B. Stanley, and J. M. Alonso. *Timing precision in population coding of natural scenes in the early visual system*. PLoS biology, 6(12), December 2008.
- [12] A. S. Ecker, P. Berens, G. A. Keliris, M. Bethge, N. K. Logothetis, and A. S. Tolias. *Decorrelated neuronal firing in cortical microcircuits*. Science, **327** (2010), 584–587.
- [13] P. Erdős and T. Gallai. *Graphs with prescribed degrees of vertices*. Mat. Lapok, **11** (1960), 264–274.
- [14] T. Hastie, R. Tibshirani, and M. Wainwright. *The Lasso:  $\ell_1$ -constrained Models in Statistics*. Manuscript, 2011.
- [15] C. Hillar, S. Lin, and A. Wibisono. *Inverses of symmetric, diagonally dominant positive matrices and applications*. <http://arxiv.org/abs/1203.6812>, 2012.
- [16] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the National Academy of Sciences (USA), **79** (1982).
- [17] J. J. Hopfield. *Pattern recognition computation using action potential timing for stimulus representation*. Nature, **376** (1995), 33–36.
- [18] E. T. Jaynes. *Information theory and statistical mechanics*. Physical Review, **106** (1957).
- [19] B. E. Kilavik, S. Roux, A. Ponce-Alvarez, J. Confais, S. Grün, and A. Riehle. *Long-term modifications in motor cortical dynamics induced by intensive practice*. The Journal of Neuroscience, **29** (2009), 12653–12663.
- [20] S. Lang. *Real and Functional Analysis*. Springer, 1993.

- [21] , W. A. Little, *The existence of persistent states in the brain*, Mathematical Biosciences, **19** (1974), 101–120.
- [22] R. C. Liu, S. Tzonev, S. Rebrik, and K. D. Miller. *Variability and information in a neural code of the cat lateral geniculate nucleus*. Journal of Neurophysiology, **86** (2001), 2789–2806.
- [23] D. M. MacKay and W. S. McCulloch. *The limiting information capacity of a neuronal link*. Bulletin of Mathematical Biology, **14** (1952), 127–135.
- [24] P. Maldonado, C. Babul, W. Singer, E. Rodriguez, D. Berger, and S. Grün. *Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images*. Journal of Neurophysiology, **100** (2008), 1523–1532.
- [25] T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan. *Maximum entropy models for antibody diversity*. Proceedings of the National Academy of Sciences (USA), **107** (2010), 5405–5410.
- [26] I. Nemenman, G. D. Lewen, W. Bialek, and R. R. van Steveninck. *Neural coding of natural stimuli: information at sub-millisecond resolution*. PLoS Computational Biology, **4** (2008), e1000025.
- [27] S. Neuenschwander and W. Singer. *Long-range synchronization of oscillatory light responses in the cat retina and lateral geniculate nucleus*. Nature, **379** (1996), 728–733.
- [28] N. Proudfoot and D. Speyer. *A broken circuit ring*. Beiträge zur Algebra und Geometrie, **47** (2006), 161–166.
- [29] F. Rieke, D. Warland, R. R. van Steveninck, W. Bialek. *Spikes: exploring the neural code*. MIT Press, 1999.
- [30] W. Russ, D. Lowery, P. Mishra, M. Yae, and R. Ranganathan. *Natural-like function in artificial WW domains*. Nature, **437** (2005), 579–583.
- [31] R. Sanyal, B. Sturmfels, C. Vinzant. *The entropic discriminant*. Advances in Mathematics, 2013.
- [32] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. *Weak pairwise correlations imply strongly correlated network states in a neural population*. Nature, **440** (2006), 1007–1012.
- [33] C. Shannon, *The mathematical theory of communication*, Bell Syst. Tech. J, **27** (1948).
- [34] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky. *The structure of multi-neuron firing patterns in primate retina*. Journal of Neuroscience, **26**(32):8254–8266, 2006.
- [35] J. Shlens, G. D. Field, J. L. Gauthier, M. Greschner, A. Sher, A. M. Litke, and E. J. Chichilnisky. *The structure of large-scale synchronized firing in primate retina*. Journal of Neuroscience, **29**(15):5022–5031, 2009.
- [36] M. Socolich, S. Lockless, W. Russ, H. Lee, K. Gardner, and R. Ranganathan. *Evolutionary information for specifying a protein fold*. Nature, **437** (2005), 512–518.
- [37] A. Tang, D. Jackson, J. Hobbs, W. Chen, J. Smith, and et al. *A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro*. Journal of Neuroscience, **28** (2008), 505–518.
- [38] G. Tkacik, E. Schneidman, M. Berry, and W. Bialek. *Ising models for networks of real neurons*. <http://arxiv.org/abs/q-bio/0611072>, 2006.

- [39] P. J. Uhlhaas, G. Pipa, G. B. Lima, L. Melloni, S. Neuenschwander, D. Nikolić, D. and W. Singer. *Neural synchrony in cortical networks: history, concept and current status*. Frontiers in Integrative Neuroscience, **3** (2009), 1–19.
- [40] J. D. Victor and K. P. Purpura. *Nature and precision of temporal coding in visual cortex: a metric-space analysis*. Journal of Neurophysiology, **76**(2):1310–1326, 1996.
- [41] M. Wainwright and M. I. Jordan. *Graphical models, exponential families, and variational inference*. Foundations and Trends in Machine Learning, **1**(1–2):1–305, January 2008.
- [42] S. Yu, D. Huang, W. Singer, and D. Nikolic. *A small world of neuronal synchrony*. Cerebral Cortex, **18** (2008), 2891–2901.